2020 Poverty Estimates for the U.S. States and Counties

Cesar A. Corral

April 26th, 2023

1 Data Description

In this project, we will analyze the Poverty-Estimates dataset, which contains data on Poverty estimates for the U.S., States, and counties, 2020. Data has been obtained from U.S. Department of Commerce, Bureau of the Census, low Area Income and Poverty Estimates (SAIPE) Program.

FIPS_code	Stabr	Area_name	POVALL_2020
0	US	United States	38371394
1000	AL	Alabama	714568
1001	AL	Autauga County	6242
1003	AL	Baldwin County	20189
1005	AL	Barbour County	5548
1007	AL	Bibb County	3549

FIPS_code	Stabr	Area_name	POVALL_2020
56035	WY	Sublette County	610
56037	WY	Sweetwater County	3187
56039	WY	Teton County	1207
56041	WY	Uinta County	1697
56043	WY	Washakie County	739
56045	WY	Weston County	679

2 Proportion of poverty for each county based on total US poverty (people)

Area_name	Stabr	proportions
Los Angeles County	CA	0.0336023
Harris County	TX	0.0194714
Cook County	IL	0.0168405
Maricopa County	AZ	0.0135725
Kings County	NY	0.0116456
Miami-Dade County	FL	0.0104191

Area_name	Stabr	proportions
Hooker County	NE	1.3e-06
Skagway Municipality	AK	1.2e-06
Kenedy County	TX	1.1e-06
King County	TX	9.0e-07
Loving County	TX	2.0e-07
Kalawao County	HI	0.0e+00

Based on these results, we can notice that counties such as Los Angeles, Harris, Cook, Maricopa, Kings, and Miami-Dade have a relatively high proportion of poverty compared to other counties. Also, counties such that Hooker, Skagway, Kenedy, King, Loving, and Kalawao have a relatively low proportion of poverty compared to other counties.

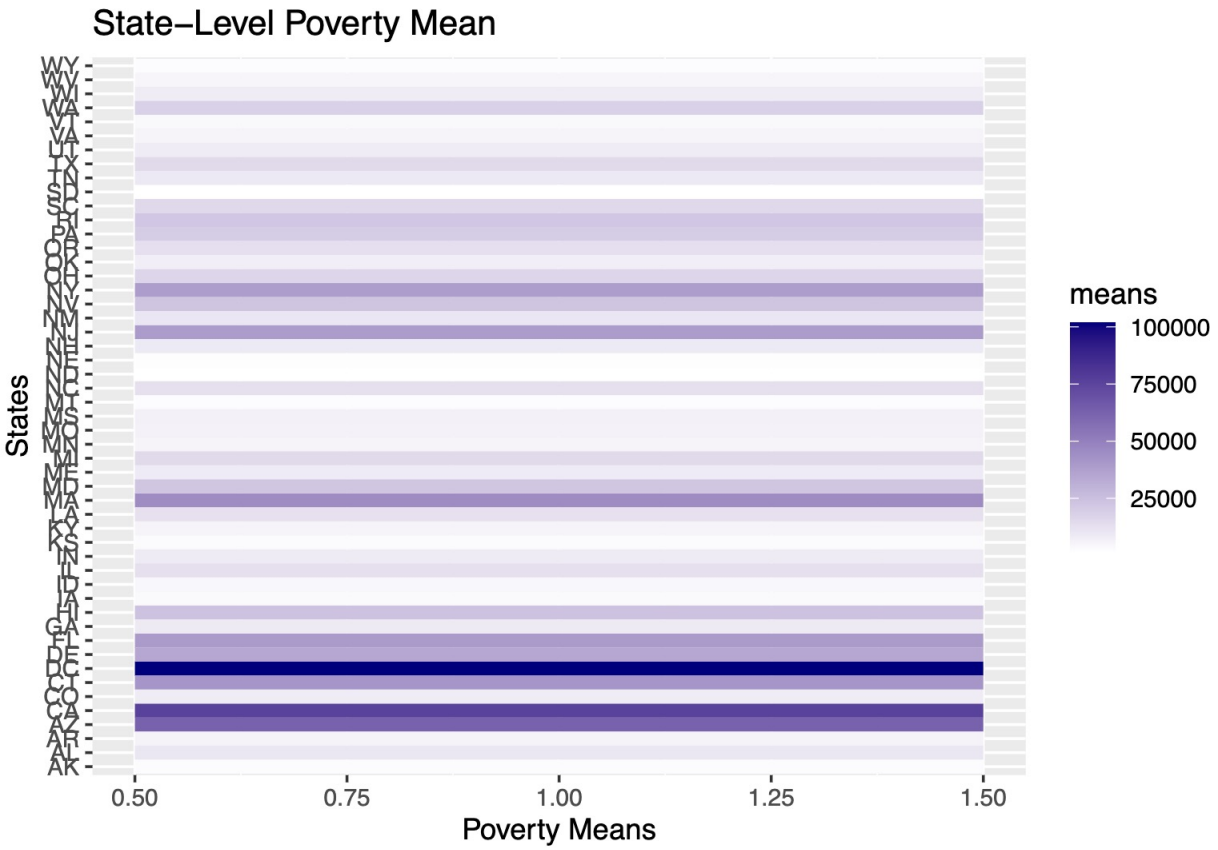
3 Poverty Mean for each state

Stabr	means
DC	101959.00
CA	76192.53
AZ	62170.20
MA	44921.21
CT	41679.38
FL	39442.43

Stabr	means
MT	2334.125
AK	2290.467
WY	2279.130
NE	1856.108
SD	1513.727
ND	1429.830

Based on these results, we can notice that states such as District of Columbia, California, Arizona, Massachusetts, Connecticut, and Florida have a relatively high mean of poverty compared to other states. Also, states such that Montana, Arkansas, Wyoming, Nebraska, South Dakota, and North Dakota have a relatively low mean of poverty compared to other states.

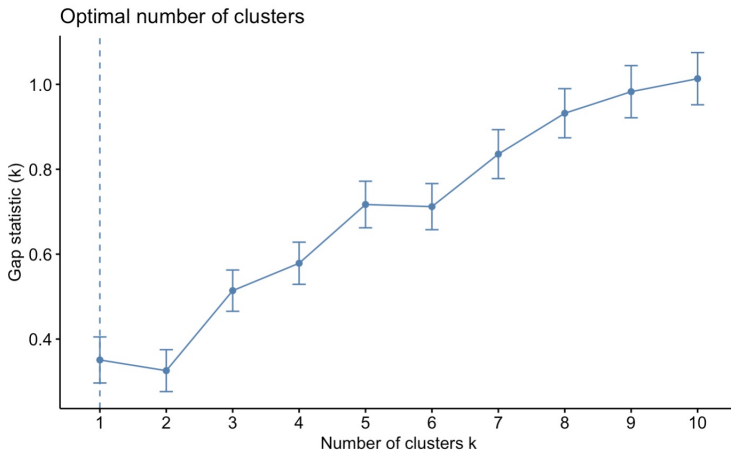
4 Heat Map for state-level (mean) poverty



Based on this result, we lead to the same conclusions like before; California is the state with the highest mean of poverty, and North Dakota, with the lowest one.

5 K-Means Clustering for states

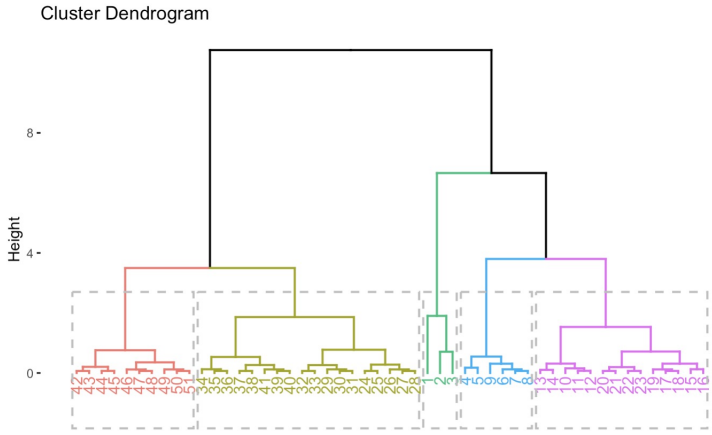
```
K-means clustering with 3 clusters of sizes 22, 20, 9

Cluster means:
       means    numbers
1 -0.6299700  0.9753756
2 -0.1166031 -0.4372373
3  1.7990447 -1.4126129

Clustering vector:
 [1] 3 3 3 3 3 3 3 3 3 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
[45] 1 1 1 1 1 1 1

Within cluster sum of squares by cluster:
[1]  4.293160  4.198145 10.664076
 (between_SS / total_SS =  80.8 %)

Available components:

 [1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
 [6] "betweenss"    "size"         "iter"         "ifault"       "clust_plot"
[11] "silinfo"      "nbclust"      "data"
[1] 19.15538
[1] 100
[1] 80.84462
[1] 0.2369407
```

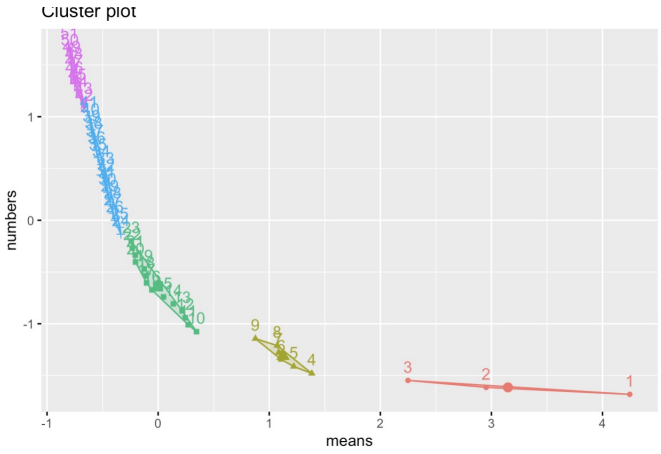


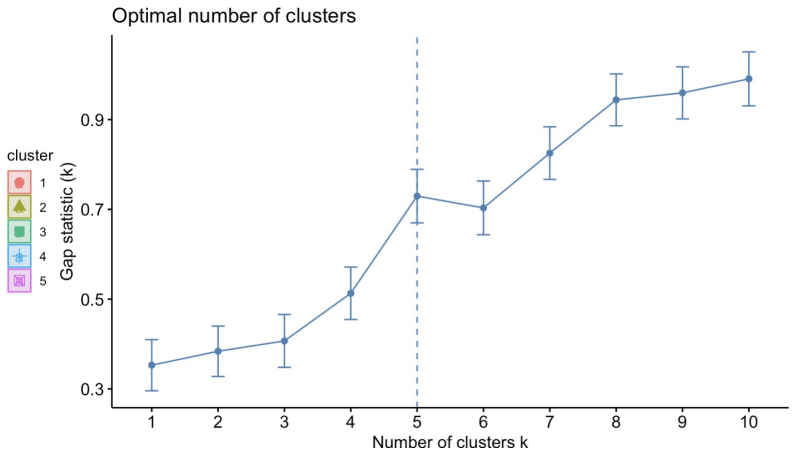
Based on these R-outputs, we can notice that, for $k = 3$, WB = 80.8%, which means that the WB ratio indicates that the clusters are relatively good separated.

6 Hierarchical Clustering for states

```
Call:
stats::hclust(d = x, method = hc_method)

Cluster method   : ward.D2
Distance         : euclidean
Number of objects: 51


            Length Class   Mode
merge       100    -none-  numeric
height       50    -none-  numeric
order        51    -none-  numeric
labels        0    -none-  NULL
method        1    -none-  character
call          3    -none-  call
dist.method   1    -none-  character
cluster      51    -none-  numeric
nbclust       1    -none-  numeric
silinfo       3    -none-  list
size          5    -none-  numeric
data        102    -none-  numeric
gap_stat      6    clusGap list
Clustering k = 1,2,..., K.max (= 10): .. done
Bootstrapping, b = 1,2,..., B (= 100)  [one "." per sample]:
.................................................. 50
```

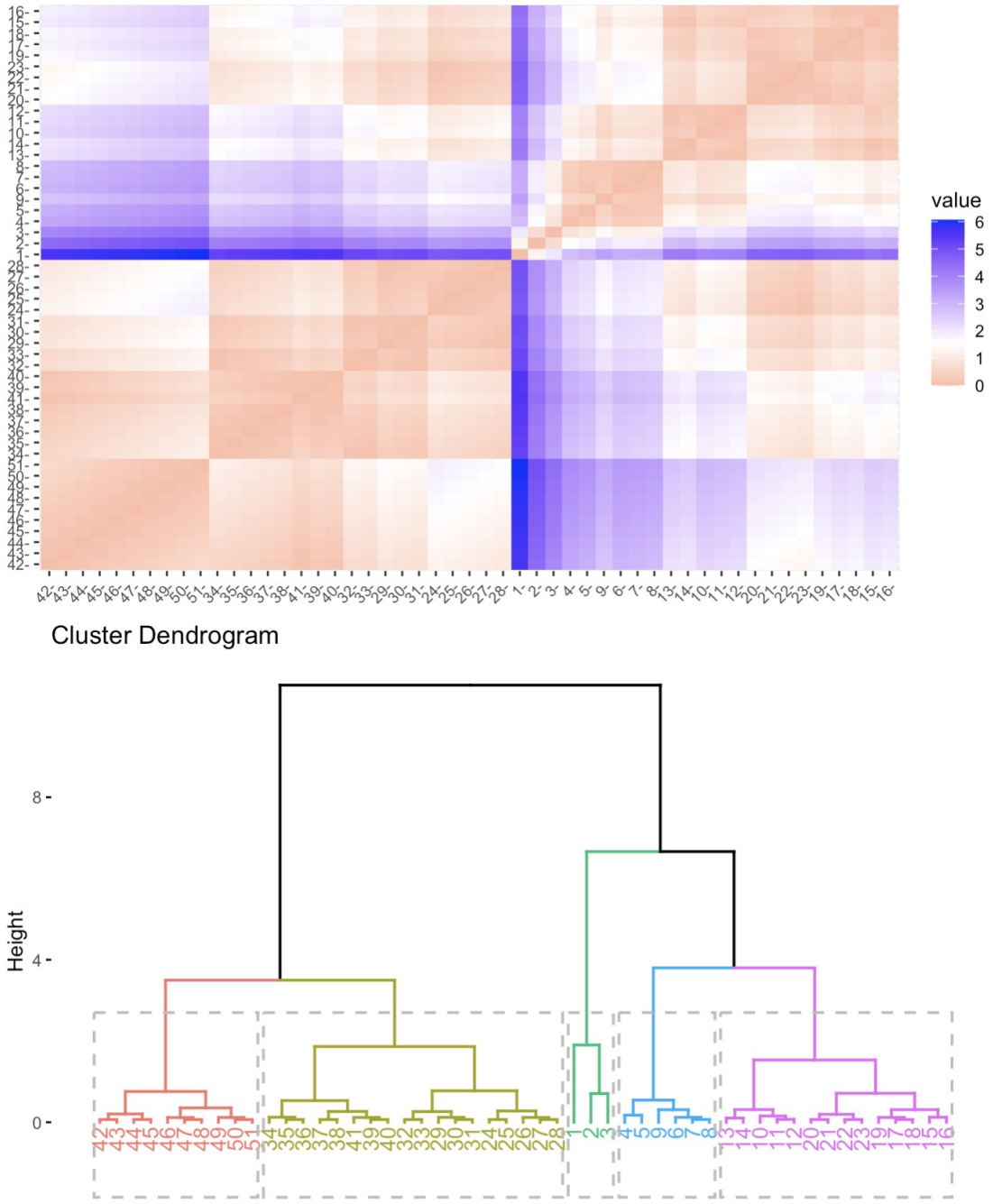






Based on these R-outputs, the Dendrogram separates the means in 5 different clusters, and it looks like they are good separated based on the Gap statistic(k).

7 Heat Maps for K-Means and Hierarchical Clustering



Based on these R-outputs, we can notice that the heat maps for K-Means and Hierarchical Clustering follow a similar pattern; however, Hierarchical Clustering map separates the means in more efficient clusters.

8 Concluding Remarks

Based on the previous results, we can conclude that Hierarchical Clustering method is has more significant results in comparison with K-Means Clustering; it leads to the conclusion that states are relatively related based on their means of poverty.